New Jersey Institute of Technology

Digital Commons @ NJIT

Fall 2024

DS 675-001, 003, 005, 101: Machine Learning

Khalid Bakhshaliyev

Follow this and additional works at: https://digitalcommons.njit.edu/ds-syllabi

New Jersey Institute of Technology
Ying Wu College of Computing
Data Science Department
Machine Learning
Code: DS675_001
Mode: Face-to-Face
Instructor: Khalid Bakhshaliyev
Office: GITC 4321A
Email: khalid.bakhshaliyev+ds675001@njit.edu
Office Hours: Wednesday 1:40 pm to 2:30pm, Friday 1pm-1:30pm

Teaching Assistant: TBD
Teaching Assistant Office Hours : TBD

Note: Please indicate **DS675_001** in your emails. Your messages will be answered by the end of the next day. Grades for all items will be getting posted during the week after their due date. For issues with your grades, contact the grader and cc the instructor.

Tutoring. NJIT provides a tutoring service. Please contact one of the available tutors. Their availability and contact information can be found here.

Course Description

This course is an introduction to machine learning and contains both theory and applications. Students will get exposure to a broad range of machine learning methods and hands-on practice on real data. Topics include Bayesian classification, perceptron, neural networks, logistic regression, support vector machines, decision trees, random forests, boosting, dimensionality reduction, unsupervised learning, regression, and learning new feature spaces. There will be several programming assignments, one course project, one mid-term and one final exam.

Machine Learning develops computer programs that can improve their performance by tapping into existing data and taking feedback from the environment. Systems based on ML have already exceeded human performance in several tasks, including image medical image classification and games like Chess and Go. ML has also made leaps in even more complicated tasks, like Natural Language Processing or self-driving vehicles, and it has even produced art that imitates the style of human artists! This course offers an intense introduction to the fundamental ML concepts and algorithms that constitute the core of these spectacular developments. It takes you on a tour from the basic mathematical notions and algorithms to some of the recent developments, e.g. Deep Networks or Recurrent Networks. You will gain exposure to cutting-edge ML development tools such as Scikit-learn and PyTorch via hands-on assignments and projects that will instill a working and immediately applicable knowledge of ML methods and will prepare you for more advanced ML courses.

Prerequisites

The course does not have other course prerequisites.

Background on some basic calculus, linear algebra, probability and programming ability is required. The following free online materials are recommended for reviewing this background:

        [Mathematics for Machine Learning](#)
        [A visual guide to NumPy](#)

Course Textbooks

There is no required course textbook. The course will draw material from several sources, including the instructor's own notes. Some optional resources include:

        A course in Machine Learning
        [Online book](#)
        Haul Daume III
        Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow:

Concepts, Tools and Techniques to Build Intelligent Systems (2nd Edition)
Aurelien Geron. (ISBN-13: 978-1492032649)

        Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2, (3rd edition)
S. Raschka, V. Mirjalili, Packt Publishing, ISBN-10: 1789955750

        The Elements of Statistical Learning, (2nd Edition)
        T. Hastie, R. Tibshirani, J. Friedman

        Introduction to Statistical Learning (https://www.statlearning.com/)
        G. James, D. Witten, T. Hastie, R. Tibshirani

Learning Outcomes

By the end of the course, you will be able to:

    a.   Identify the main types of Machine Learning (ML)
    b.   Evaluate the quality of online resources related to ML
    c.   Recognize problems amenable to ML methods
    d.   Describe and explain a wide variety of ML algorithms
    e.   Apply various ML algorithms in novel situations
    f.   Evaluate the performance of ML models
    g.   Modify ML models in order to improve their performance
    h.   Adapt ML algorithms and models to the given data and application

---

Coursework, Assessment and Related Outcomes

Hands-On Assignments [25%]. Six hands-on assignments of equal grading weight. The weakest of the six grades is dropped from the calculation. [Outcomes: c,e-h]

Quizzes [15%]. Weekly paper/Canvas quizzes reinforcing the material of each module. For canvas quizzes no browser restrictions, **open** books/notes/web. [Outcomes: a,d]

Midterm exam [25%]. In-person exam, 120 minutes. Each student is allowed to bring at most 4 pages (A4 or Letter) of notes. [Outcomes: a,d,e]

Final exam [25%]. Cumulative, 120 minutes. Otherwise, similar to midterm exams. [Outcomes: a,d,e]

Project [10%]. The project will consist of two to three milestones:

Participate in Weekly Discussion [extra 10%]

Grading Scheme and Letter Grades. The conversion of raw total scores to letter grades will be based on grouping the scores into clusters and then assigning a letter grade to each cluster. Projected letter grades will be communicated after the midterm exam and will be constantly updated throughout the rest of the semester, to reflect the underlying clusters. The final letter grade assignment will always be in accordance with the graduate grade legend.

Grading Feedback. Assignment marks will be accompanied with solutions and general feedback summarizing common mistakes. Individual grading feedback will be given whenever possible. Further clarifications can be provided via direct communication with the instructor and the course grader.

Late Work Policy. In the case when a student is unable to complete an assignment or other serious reasons, these must be communicated and documented promptly. In any other case, each hour of delay after the due date will incur a 2% score reduction. No extensions will be granted. However, the lowest programming score and the two lowest quiz assignments scores for each student will be dropped.

Course Topic Schedule

Week 1	1. Introduction 2. Nearest neighbors and key notions
Week 2	3. Linear regression, polynomial regression, feature engineering
Week 3	4. Linear separability, perceptron
Week 4	5. The sigmoid neuron, logistic regression 6. Regularization
Week 5	7. Support Vector Machines, Kernels 8. Decision Trees, Random Forests
Week 6	9. Validation and Hyperparameter Tuning 10. Feature Selection
Week 7	11. Ensembles and Boosting
Week 8	12. A probabilistic perspective (midterm week)
Week 9	13. Dimensionality reduction, Kernel PCA 14. Unsupervised Clustering

Week 10	15. Introduction to Artificial Neural Networks 16. MLP classifiers
Week 11	17. Autoencoders 18. Computation with GPUs, Automatic differentiation
Week 12	19. Convolutional Neural Networks
Week 13	20. Recurrent Neural Networks
Week 14	21. Selected project presentations

Course Policies

Email

Use of your NJIT email or Canvas inbox is strongly encouraged. Include your class section in the email.

Grade Corrections

Check the grades in course work and report errors promptly. Please try and resolve any issue within one week of the grade notification.

Exam and Proctoring Policy

See the NJIT Online Course Exam Proctoring page for information on proctoring tools and requirements.

Incomplete

A grade of I (incomplete) is given in rare cases where work cannot be completed during the semester due to documented long-term illness or unexpected absence for other serious reasons. A student needs to be in good standing (i.e. passing the course before the absence) and receives a provisional **I** if there is no time to make up for the documented lost time; an email with a timeline of what is needed to be done will be sent to the student. Note that an **I** must always be resolved by the end of the next semester.

Collaboration and External Resources for Assignments

Some homework problems will be challenging. You are advised to first try and solve all the problems **on your own**. For problems that persist you are welcome to talk to the course assistant or the instructor. You are also allowed to collaborate with your classmates and search for solutions online. But you should use such solutions only if you understand them completely (admitting that you don't understand something is way better than copying things you don't understand). Also make sure to give the appropriate credit and citation.

Requesting Accommodations

If you need an accommodation due to a disability please contact Scott Janz, Associate Director of the Office of Accessibility Resources and Services, Kupfrian Hall 201 to discuss your specific needs. A Letter of Accommodation Eligibility from the office authorizing student accommodations is required.

NJIT Services for Students, Including Technical Support

Please follow this link.

Canvas Accessibility Statement
Please follow this link.
Academic Integrity
Academic Integrity is the cornerstone of higher education and is central to the ideals of this course and the university. Cheating is strictly prohibited and devalues the degree that you are working on. As a member of the NJIT community, it is your responsibility to protect your educational investment by knowing and following the academic code of integrity policy that is found a this link.

Please note that it is my professional obligation and responsibility to report any academic misconduct to the Dean of Students Office. Any student found in violation of the code by cheating, plagiarizing or using any online software inappropriately will result in disciplinary action. This may include a failing grade of F, and/or suspension or dismissal from the university. If you have any questions about the code of Academic Integrity, please contact the Dean of Students Office at dos@njit.edu